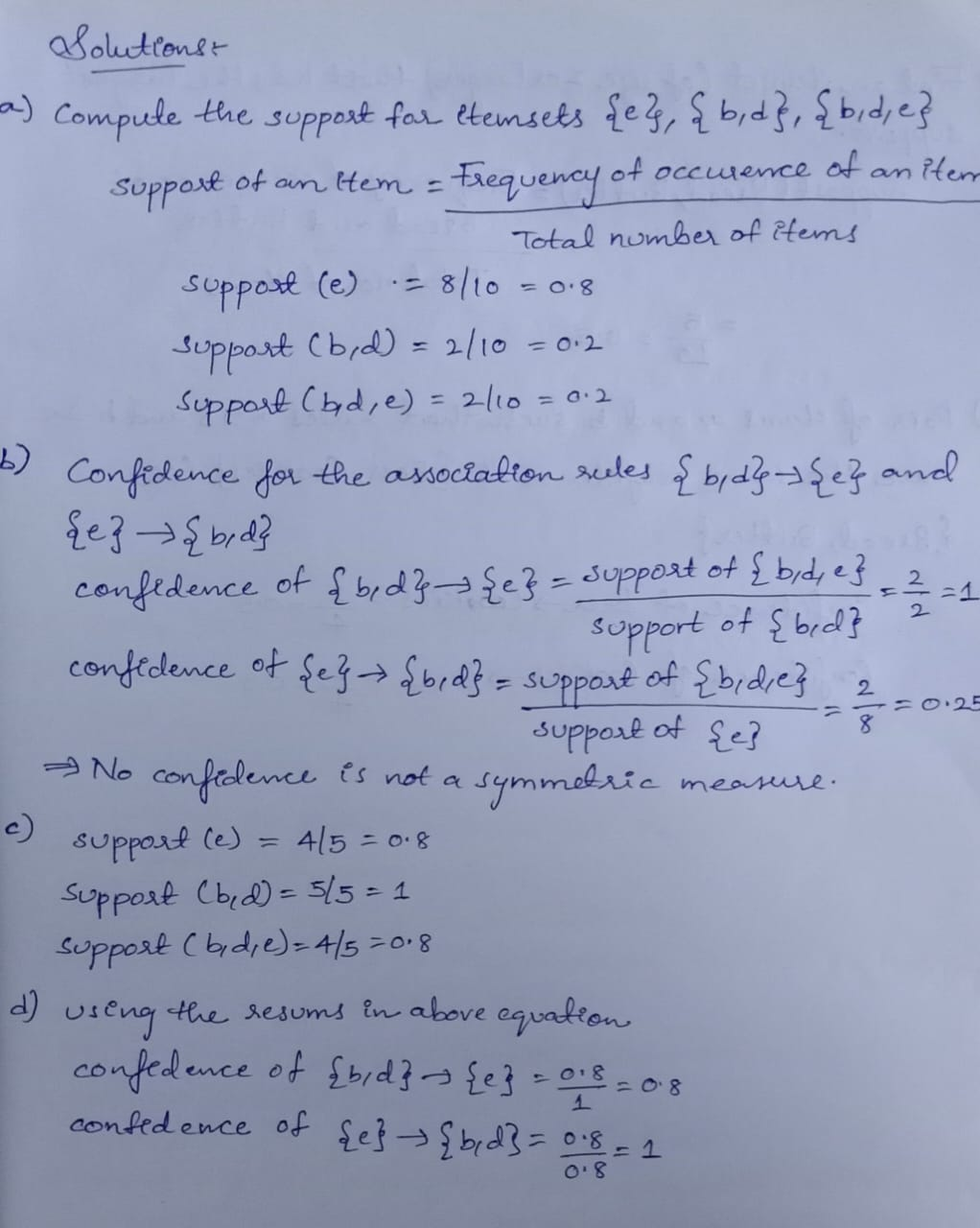
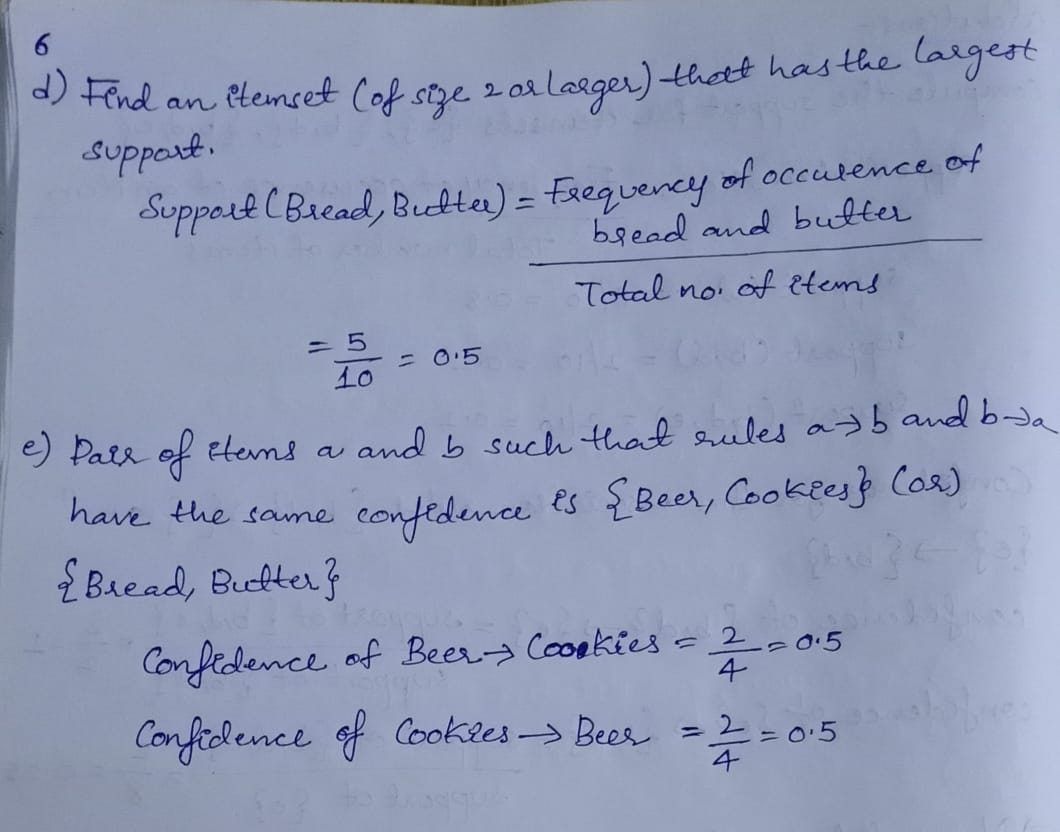
Data Mining Assignment 3

1) Read Chapter 6 (only sections 6.1 and 6.7).  
  
2) Do Chapter 6 textbook problem #2 (parts a,b,c,d only) on page 404.



3) Do Chapter 6 textbook problem #6 (parts d,e only) on page 406.

  
  
4) Using the data at [www.stats202.com/more\_stats202\_logs.txt](http://www.stats202.com/more_stats202_logs.txt) and treating each row as a "market basket" compute the support and confidence for the rule ip=65.57.245.11 → "Mozilla/5.0 (X11; U; Linux i686 (x86\_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3".

State what the support and confidence values mean in plain English in this context.

Answer:

1. Support represents the popularity of that product of all the product transactions. Support of the product is calculated as the ratio of the number of transactions includes that product and the total number of transactions.
2. Support of the product = (Number of transactions includes that product)/ (Total number of transactions)
3. Confidence can be interpreted as the likelihood of purchasing both the products A and B. Confidence is calculated as the number of transactions that include both A and B divided by the number of transactions includes only product A.
4. Confidence (A=>B) = (Number of transactions includes both A and B)/ (Number of transactions includes only product A)